Auction Price Prediction & Image Features

Presented by: Yifei Wang, Bingying (Echo) Liu

Auction Process

Item on sale

Bidding

Highest price

Buyer's premium

Project goals

Price prediction

Interpretability

purple wave auction®

Straight. Simple. Sold.®

Advertising

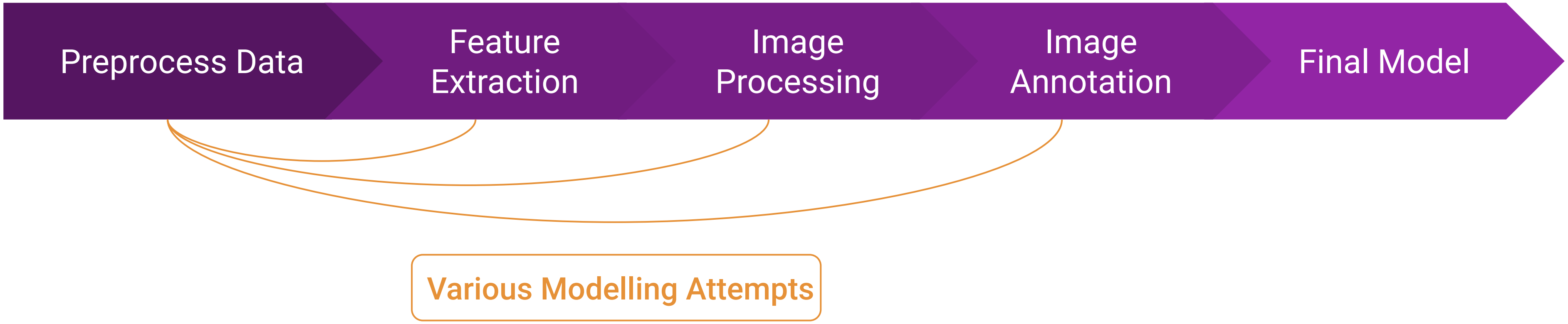
Auction arrangement

Increase expected price

Sellers

Estimate vehicle value

Fix or not fix


Preprocess Data
Feature Extraction
Image Processing
Image Annotation
Final Model
Various Modelling Attempts

Timeline

Preprocess Data

Tabular Data

- Final auction price
- Make, Year, Mileage, etc.

Text Data

- Tire, bucket, engine, etc.

Extract Features!

Non-annotated

Feature: Colorfulness Scores

10 Most Colorful

10 Least Colorful

Feature: Text Sentiment

VADER

- ❖ Lexicon and rule-based sentiment analysis tool
- ❖ Preceding tri-grams (observe 90% negation flip); Conjunctions

Engine (more than 100 characters' comments)	**VADER Sentiment Scores**
Engine access door latch area damaged, engine knocks, engine shroud damaged, new holland 450nc 5.01 four cylinder diesel engine	{'neg': 0.254, 'neu': 0.746, 'pos': 0.0, 'compound': -0.70}
81 hp, approximately one hour on replaced engine, new holland 445m2 four cylinder diesel engine, recently replaced engine	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
70 hp, john deere 5030ht014 3.01 five cylinder turbo diesel engine, one engine cover side panel missing	{'neg': 0.121, 'neu': 0.879, 'pos': 0.0, 'compound': -0.296}
Engine compartment fire damage, john deere pe5030 four cylinder turbo diesel engine, lift arm cylinder pins removed, but included	{'neg': 0.183, 'neu': 0.817, 'pos': 0.0, 'compound': -0.4215}

Image Annotation

View instructions

Instructions: Given an image, answer the following questions.

- If your answer to the first question is "Yes", a follow-up question will pop out. Otherwise, please continue to the second question.
- Please click on **"View instructions"** before moving on to your task.

1. **Does this equipment's bucket contain rust?**

   Yes No This equipment has no bucket.

2. **Use a scale of 1 to 10 to describe how much rust the body of equipment has (please only take in account the body, not the bucket).**

   1 means 'barely any rust' and 10 means 'rust all over the body'.

   1 2 3 4 5 6 7 8 9 10

Submit

Sample 100 images

13 questions/image

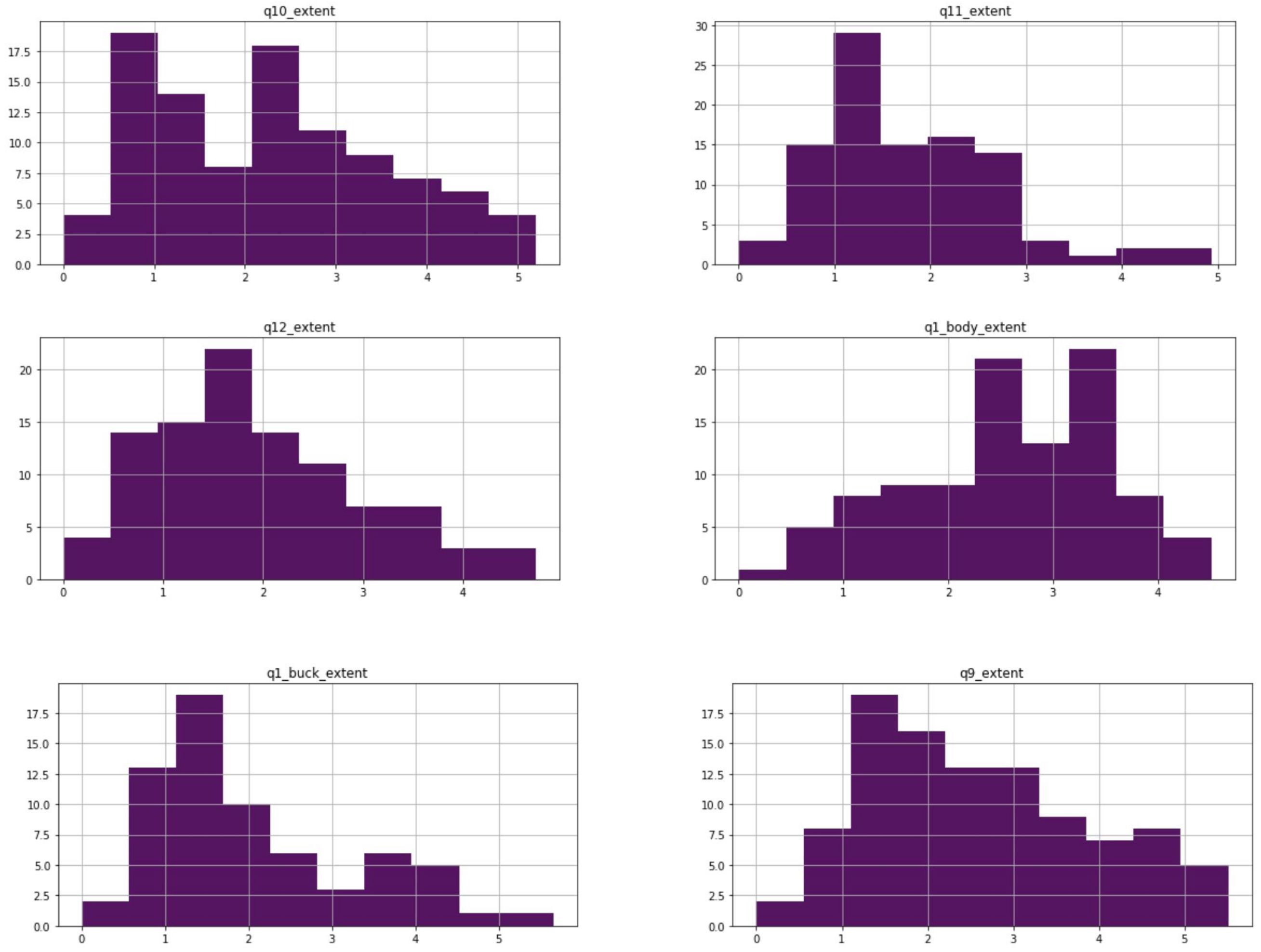
3 workers/question

$ 83.94 in total

13 questions

Yes-no & Quantitative

Background
Rust extent
Brightness of color
Dirt level



Distribution of Standard Deviations of All Numerical Answers for Each Image

Annotation Quality:

- ❖ Except for body rust extent, all other annotations are within 2-3 std dev away.

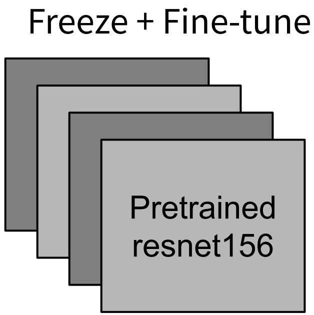
Final Model

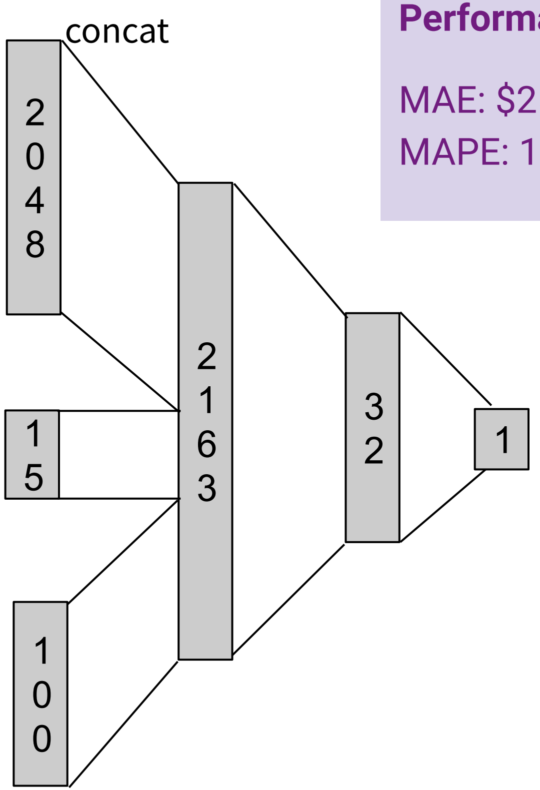
Performance:

MAE: $2541.85

MAPE: 19.06%

Image





Tabular (+ Annotation)

Year
Make
Mileage
Colorfulness
etc.

Text

…, new factory remanufactured head, new ...

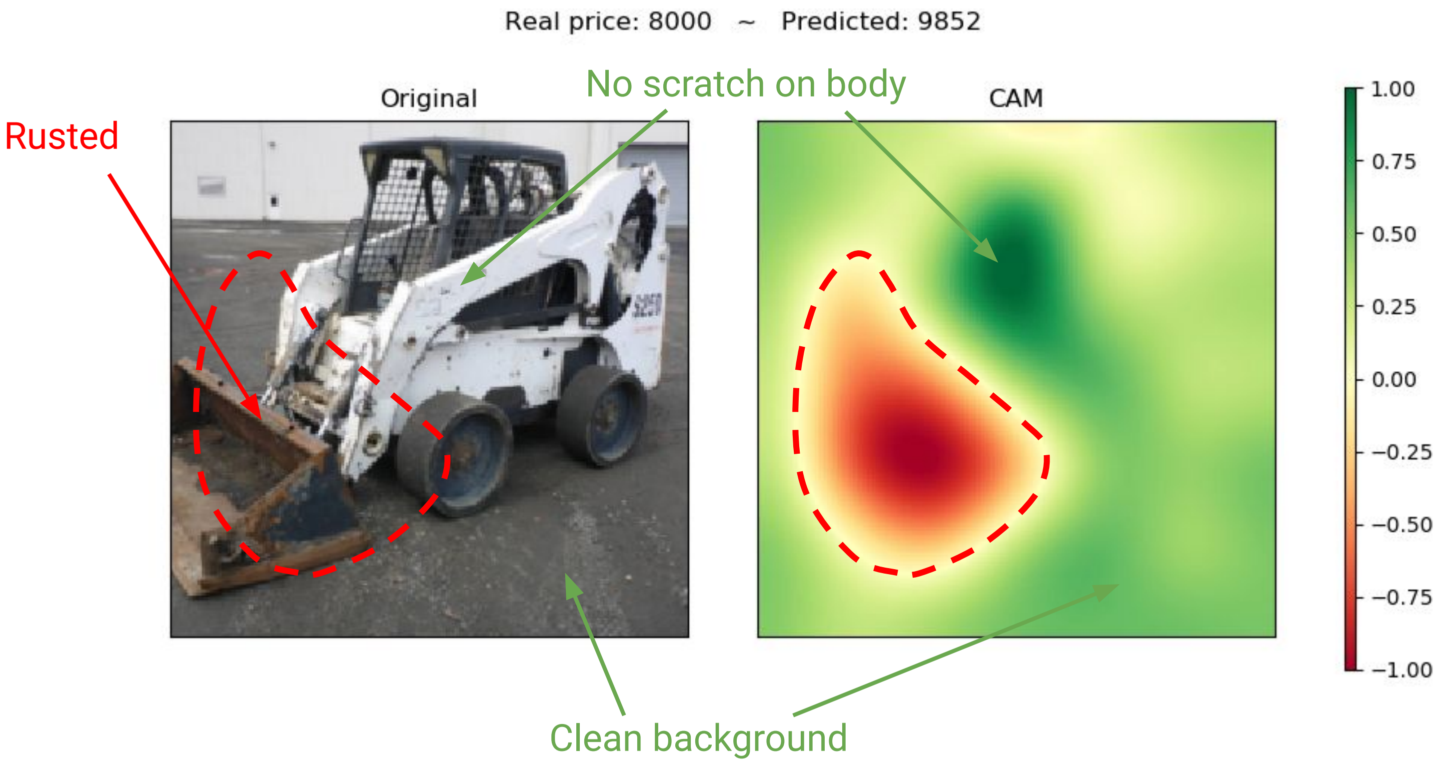
Price

Interpretability - Images

- Hard to generate quantitative interpretation on images
  - without enough human annotations

- Try to qualitatively interpret images
  - Indicate which part of the image is driving up/down the price

Interpretability - Attention Map



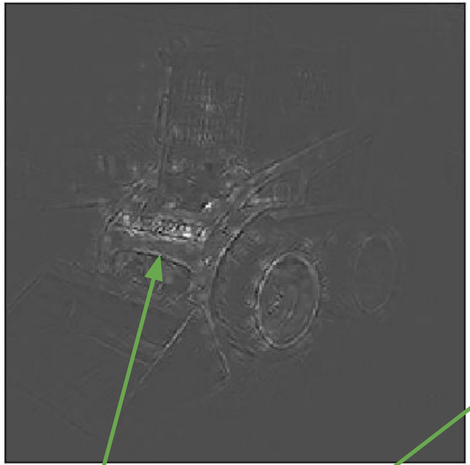
Interpretability - Saliency Map

Real price: 15000 ~ Predicted: 11249

Original

$$\frac{\partial\, \text{price}}{\partial\, \text{image}} =$$

Monochromatic



Colorful

Clear brand logo and clean body

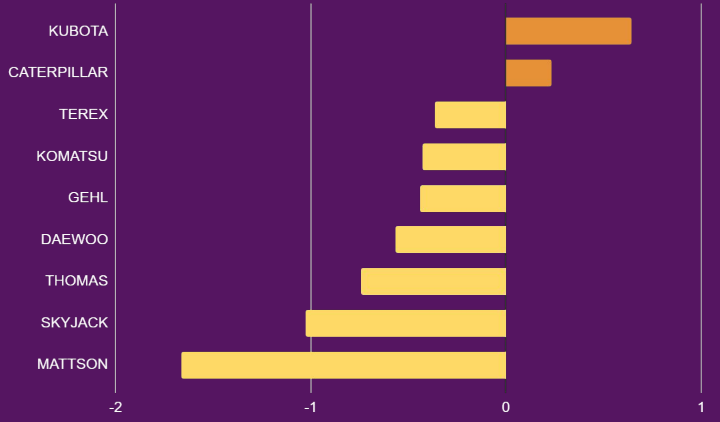
Model Comparisons

Model	Parameters	Sample Size	R^2	MAE
MLR	Hours_final, age_at_sale	6167	0.392	4076
MLR	Hours_final, age_at_sale, **parts_sentiment**	6167	0.395	4062
MLR	Hours_final, age_at_sale, **colorfulness_score**	6167	0.396	4047
MLR	Hours_final, age_at_sale, **month_of_sold_date**	6167	0.401	4066
MLR	Hours_final, age_at_sale, **make**	6167	0.434	3926
MLR	Hours_final, age_at_sale, **make, month_of_sold_date, colorfulness_score, parts_sentiment**	6167	0.442	3902
H2O_DRF	Hours_final, age_at_sale, **make, month_of_sold_date, colorfulness_score, parts_sentiment**	6167	0.489 (MSE)	3391
MLR	Hours_final, age_at_sale	96	0.431	5384
MLR	Hours_final, age_at_sale, **make**	96	0.570	4998
MLR	Hours_final, age_at_sale, **make, mturk**	96	0.633	Unstable Because of Overfitting
MLR	Hours_final, age_at_sale, **make, mturk, parts_sentiment, colorfulness**	96	0.676	
MLR	Hours_final, age_at_sale, **make, mturk, parts_sentiment, colorfulness , month_sold**	96	0.802	
H2O_DRF	Hours_final, age_at_sale, **make, mturk, parts_sentiment, colorfulness , month_sold**	96	0.735 (MSE)	

Brand Premium

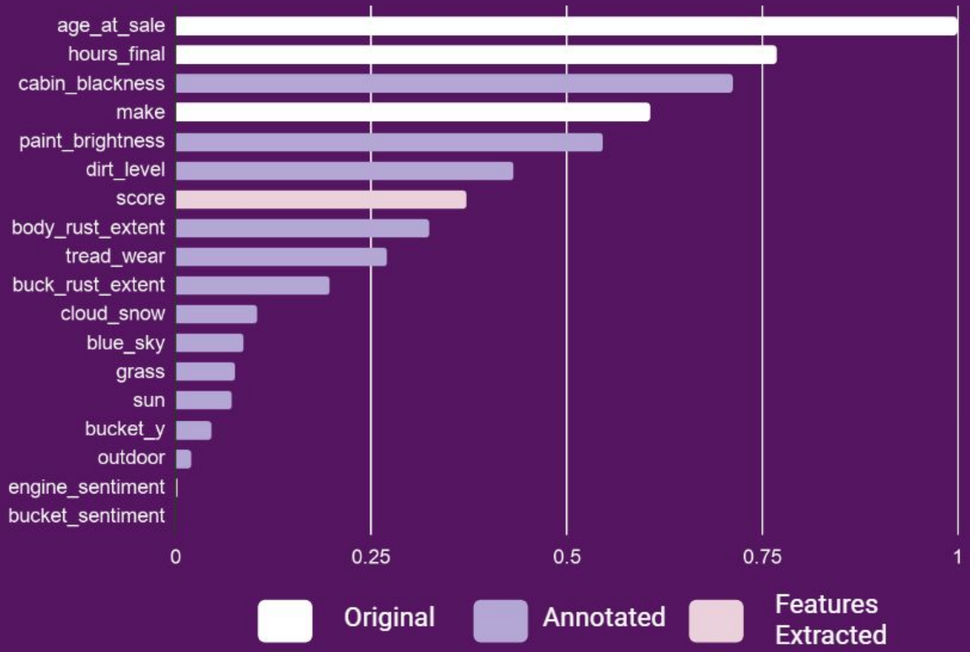
Final selling price is on log scale

** chart is based on coefficients of MLR*



Top Feature Importance

** Chart is based on distributed random forest*



Seasonality Influence on Sales Price

- ❖ Oct, Apr: Positive influence
- ❖ Jan, Sep: Negative influence

Quantitative Interpretation

https://skidsteer-interpret.herokuapp.com/

Interactive Visualizations

Summary

- ❖ Interpretable features (machine and human)
- ❖ Price prediction

Recommendations

- ❖ Image annotation (MTurk)
- ❖ Vehicle cleanliness

Thank you!

Yifei Wang & Bingying Liu

II. Feature Extraction - Image Colorfulness

- Haslera and Susstrunk (2003)
- They ask people to rate images using 7 categories of colorfulness
- Propose a metric that correlates to 95.3% of the experiment data
- ***Opponent color space representation***

$$Rg = R - G$$

$$Yb = (1/2)\,(R + G) - B$$

$$\sigma_{rgyb} = \sqrt{{\sigma^2}_{rg} + {\sigma^2}_{yb}}$$

$$\mu_{rgyb} = \sqrt{{\mu^2}_{rg} + {\mu^2}_{yb}}$$

$$C = \sigma_{rgyb} + 0.3 * \mu_{rgyb}$$